

# PRINTABLE REPRESENTATIONS FOR TIME-BASED MEDIA

## Inventors:

Jonathan J. Hull  
Jamey Graham  
Peter E. Hart

## CROSS-REFERENCES TO RELATED APPLICATIONS

[0001] This application claims the benefit of the following provisional patent applications, each of which is incorporated by reference in its entirety: U.S. Provisional patent application entitled “Printer Including One or More Specialized Hardware Devices” filed on September 25, 2003, having serial number 60/506,303; U.S. Provisional patent application entitled “Printer For Non-Paper and Multiple Media Types” filed on September 25, 2003, having serial number 60/506,411; and U.S. Provisional patent application entitled “Systems and Methods for Printing Audio or Video Paper” filed on September 25, 2003, having serial number 60/506,263.

[0002] This application is a continuation-in-part of the following co-pending U.S. Patent Applications (hereinafter referred to as the “Video Paper Applications”), each of which is hereby incorporated by reference in its entirety: U.S. Patent Application Serial No. 10/001,895, “Paper-based Interface for Multimedia Information,” filed November 19, 2001; U.S. Patent Application Serial No. 10/001,849, “Techniques for Annotating Multimedia Information,” filed November 19, 2001; Application Serial No. 10/001,893, “Techniques for Generating a Coversheet for a paper-based Interface for Multimedia Information,” filed November 19, 2001; U.S. Patent Application Serial No. 10/001,894,

“Techniques for Retrieving Multimedia Information Using a Paper-Based Interface,” filed November 19, 2001; U.S. Patent Application Serial No. 10/001,891, “Paper-based Interface for Multimedia Information Stored by Multiple Multimedia Documents,” filed November 19, 2001; U.S. Patent Application Serial No. 10/175,540, “Device for Generating a Multimedia Paper Document,” filed June 18, 2002; and U.S. Patent Application Serial No. 10/645,821, “Paper-Based Interface for Specifying Ranges,” filed August 20, 2003.

[0003] This application is also related to the following co-pending patent applications, each of which is hereby incorporated by reference in its entirety: U.S. patent application entitled, “Printer Having Embedded Functionality for Printing Time-Based Media,” to Hart et al., filed March 30, 2004, Attorney Docket No. 20412-08340; U.S. patent application entitled, “Printer With Hardware and Software Interfaces for Peripheral Devices,” to Hart et al., filed March 30, 2004, Attorney Docket No. 20412-08383; U.S. patent application entitled, “Printer User Interface,” to Hart et al., filed March 30, 2004, Attorney Docket No. 20412-08455; and U.S. Patent Application entitled “Networked Printing System Having Embedded Functionality for Printing Time-Based Media,” to Hart, et al., filed March 30, 2004, Attorney Docket Number 20412-8341.

## **BACKGROUND OF THE INVENTION**

### **1. Field of the Invention**

[0004] The present invention relates to systems and methods for generating printable representations for time-based media.

## 2. Description of the Background Art

[0005] Conventional printers are currently used to generate documents of various different formats and based upon different types of content. However, while conventional printers can produce images of text and/or pictures, conventional printers are limited in their ability to effectively generate representations of multimedia content. Conventional printers print onto a fixed medium, such as paper, and thus they are unable to effectively capture the elements of time-based media.

[0006] Yet, the capability to easily review time-based media content is commonly needed today. To search for desired features within time-based media content currently, one must actually review the content itself, skimming to find the desired information. For example, a user may have to manually skim an audio recording of a radio talk show to find content on a particular topic or to find discussions by a particular speaker. Due to these limitations in conventional printers, there is currently no easy way for users to search through a lengthy media segment to identify and extract particular features of interest from the media content. Additionally, there is no way for users to create an easily readable representation of media that provides useful information about the media.

[0007] Moreover, media content is commonly only available in digital form. However, for many users, a digital format is not the optimal format in which to view information. While viewing media information in digital form is adequate for some users, many users find it easier to comprehend and assimilate information when the information is printed on a paper medium. Nonetheless, there is not currently available a

mechanism for generating a paper-based representation of time-based media through which the user can review or even access media content.

**[0008]** Therefore, what is needed is a system and methods for generating a representation of time-based media that can be paper-based and can provide users with the ability to extract defined features in the multimedia content.

### **SUMMARY OF THE INVENTION**

**[0001]** The present invention overcomes the deficiencies and limitations of the prior art with a system and method for generating a representation of time-based media. The system of the present invention includes a feature extraction module for extracting features from media content. For example, the feature extraction module can detect solos in a musical performance, or can detect music, applause, speech, and the like. A formatting module formats a media representation generated by the system. The formatting module also applies feature extraction information to the representation, and formats the representation according to a representation specification. In addition, the system can include an augmented output device that generates a media representation based on the feature extraction information and the representation specification. The representation can be generated in a paper-based format, in digital format, or in any other representation formats. The representation generated can include user-selectable identifiers that enable random access to points along a media content timeline.

**[0002]** The methods of the present invention include extracting features from media content, and formatting a media representation using the extracted features and

based on a specification or data structure specifying the representation format. The methods can also include generating a media representation based on the results of the formatting.

### **BRIEF DESCRIPTION OF THE DRAWINGS**

[0003] The invention is illustrated by way of example, and not by way of limitation in the figures of the accompanying drawings in which like reference numerals refer to similar elements.

[0004] **Figure 1** is a block diagram of a system for generating a representation of the multimedia data.

[0005] **Figure 2** is a block diagram of an exemplary architecture for one embodiment of the system of Figure 1.

[0006] **Figure 3** is a flowchart of a method for operation of an audio paper production system.

[0007] **Figure 4** is a flowchart of a method for operation of a formatting module in formatting multimedia content.

[0008] **Figure 5** is a flowchart of a method of generating barcodes for a multimedia representation.

[0009] **Figure 6a** is a representation of an exemplary document format specification and audio feature extraction.

[0010] **Figure 6b** is a graphical representation of a multimedia representation generated based on the specification depicted in Figure 6a.

- [0011]        **Figure 7a** is a representation of an exemplary document format specification and audio feature extraction that includes musical solo extraction.
- [0012]        **Figure 7b** is a graphical representation of a multimedia representation generated based on the specification depicted in Figure 7a.
- [0013]        **Figure 8a** is a representation of an exemplary document format specification and audio feature extraction for a radio program.
- [0014]        **Figure 8b** is a graphical representation of a multimedia representation generated based on the specification depicted in Figure 8a.
- [0015]        **Figure 9a** is a representation of an exemplary document format specification and audio feature extraction including keywords.
- [0016]        **Figure 9b** is a graphical representation of a multimedia representation generated based on the specification depicted in Figure 9a.
- [0017]        **Figure 10a** is a representation of an exemplary document format specification and audio feature extraction for speech recognition and word searching.
- [0018]        **Figure 10b** is a graphical representation of a multimedia representation generated based on the specification depicted in Figure 10a.
- [0019]        **Figure 11a** is a representation of an exemplary document format specification and audio feature extraction for applause detection.
- [0020]        **Figure 11b** is a graphical representation of a multimedia representation generated based on the specification depicted in Figure 11a.
- [0021]        **Figure 12a** is a representation of an exemplary document format specification and audio feature extraction for music detection.

[0022]        **Figure 12b** is a graphical representation of a multimedia representation generated based on the specification depicted in Figure 12a.

### **DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS**

[0023]        A system and method for generating a representation of time-based media is described. According to an embodiment of the present invention, a printer generates a representation of time-based media that can incorporate feature extraction information and can be formatted according to a representation specification. More specifically, the printer incorporates a format specification, feature extraction, and a formatting algorithm to produce documents that provide a visual representation for multimedia information (e.g., an audio recording), and provide an index that enables random access to points in a multimedia recording.

[0024]        For the purposes of this invention, the terms “media,” “multimedia,” “multimedia content,” “multimedia data,” or “multimedia information” refer to any one of or a combination of text information, graphics information, animation information, sound (audio) information, video information, slides information, whiteboard images information, and other types of information. For example, a video recording of a television broadcast may comprise video information and audio information. In certain instances the video recording may also comprise close-captioned (CC) text information, which comprises material related to the video information, and in many cases, is an exact representation of the speech contained in the audio portions of the video recording. Multimedia information is also used to refer to information comprising one or more objects wherein the objects include information of different types. For example,

multimedia objects included in multimedia information may comprise text information, graphics information, animation information, sound (audio) information, video information, slides information, whiteboard images information, and other types of information.

[0025] For the purposes of this invention, the terms “print” or “printing,” when referring to printing onto some type of medium, are intended to include printing, writing, drawing, imprinting, embossing, generating in digital format, and other types of generation of a data representation. Also for purposes of this invention, the output generated by the system will be referred to as a “media representation,” a “multimedia document,” a “multimedia representation,” a “document,” a “paper document,” or either “video paper” or “audio paper.” While the words “document” and “paper” are referred to in these terms, output of the system in the present invention is not limited to such a physical medium, like a paper medium. Instead, the above terms can refer to any output that is fixed in a tangible medium. In some embodiments, the output of the system of the present invention can be a representation of multimedia content printed on a physical paper document. In paper format, the multimedia document takes advantage of the high resolution and portability of paper and provides a readable representation of the multimedia information. According to the teachings of the present invention, a multimedia document may also be used to select, retrieve, and access the multimedia information. In other embodiments, the output of the system can exist in digital format or some other tangible medium. In addition, the output of the present invention can refer to any storage unit (e.g., a file) that stores multimedia information in digital format.

Various different formats may be used to store the multimedia information. These



formats include various MPEG formats (e.g., MPEG 1, MPEG 2, MPEG 4, MPEG 7, etc.), MP3 format, SMIL format, HTML+TIME format, WMF (Windows Media Format), RM (Real Media) format, Quicktime format, Shockwave format, various streaming media formats, formats being developed by the engineering community, proprietary and customary formats, and others.

**[0026]**       Reference in the specification to “one embodiment” or “an embodiment” means that a particular feature, structure, or characteristic described in connection with the embodiment is included in at least one embodiment of the invention. The appearances of the phrase “in one embodiment” in various places in the specification are not necessarily all referring to the same embodiment.

**[0027]**       In the following description, for purposes of explanation, numerous specific details are set forth in order to provide a thorough understanding of the invention. It will be apparent, however, to one skilled in the art that the invention can be practiced without these specific details. In other instances, structures and devices are shown in block diagram form in order to avoid obscuring the invention. For example, the present invention is described primarily with reference to audio content, and the representation generated by the printer is often referred to as audio paper. However, the features of the present invention apply to any type of media content and refer to media representations in formats other than paper-based formats, even if the description discusses the features only in reference to audio content and audio paper.

**[0028]**       Referring now to Figure 1, an exemplary system 100 for generating a representation of time-based media is shown. In this embodiment, there is shown an augmented output device or a printer 102 for generating multimedia representations. The

printer 102 comprises a number of components, including the following: a conventional printer 103, an audio paper production system (APPS) 108, and processing logic 106 for a printer console and for a printer driver interface.

**[0029]** The printer 102 receives multimedia data, such as audio data, and this content may be stored in a multimedia document that is accessible to system 100. The multimedia content may be stored directly on system 100, or it may be information stored on an external storage device or a server (not shown) that can be accessed by system 100. In other embodiments, instead of accessing a multimedia document, the system 100 may receive a stream of multimedia information (e.g., a streaming media signal, a cable signal, etc.) from a multimedia information source. Examples of sources that can provide multimedia information to system 100 include a television, a television broadcast receiver, a cable receiver, a video recorder, a digital video recorder, a personal digital assistant (PDA), or the like. For example, the source of multimedia information may be embodied as a radio that is configured to receive multimedia broadcast signals and to transmit the signals to system 100. In this example, the information source may be a radio antenna providing live radio broadcast feed information to system 100. The information source may also be a device such as a video recorder/player, a digital video disc (DVD) player, a compact disc (CD) player, etc. providing recorded video and/or audio stream to system 100. In alternative embodiments, the source of information may be a presentation or meeting recorder device that is capable of providing a stream of the captured presentation or meeting information to system 100. Additionally, the source of multimedia information may be a receiver (e.g., a satellite dish or a cable receiver) that is configured to capture or receive (e.g., via a wireless link) multimedia information from

an external source and then provide the captured multimedia information to system 100 for further processing.

**[0030]** Multimedia content can originate from some type proprietary or customized multimedia player, such as RealPlayer™, Microsoft Windows Media Player, and the like. In alternative embodiments, system 100 may be configured to intercept multimedia information signals received by a multimedia information source. System 100 may receive the multimedia information directly from a multimedia information source or may alternatively receive the information via a communication network (not shown).

**[0031]** Referring again to the components of printer 102, there is shown in Figure 1 a conventional printer 103 component of printer 102. The conventional printer 103 component of the printer 102 can include all or some of the capabilities of a standard or conventional printing device, such as an inkjet printer, a laser printer, or other printing device. Thus, conventional printer 102 has the functionality to print paper documents, and may also have the capabilities of a fax machine, a copy machine, and other devices for generating physical documents. More information about printing systems is provided in the U.S. Patent Application entitled “Networked Printing System Having Embedded Functionality for Printing Time-Based Media,” to Hart, et al., filed March 30, 2004, Attorney Docket Number 20412-8341, which is hereby incorporated by reference in its entirety.

**[0032]** In Figure 1, there is also shown an audio paper production system (APPS) 108 in this embodiment of the present invention. This system is referred to as an audio paper production system, but it could alternatively be a video paper production system in

other embodiments, or any other type of multimedia production system. Additionally, though the APPS 108 refers to the word “paper” in its title, the APPS 108 can also be used to generate multimedia representations in digital format and other types of formats.

**[0033]** The APPS 108 is shown in Figure 1 as being part of the printer 102. However, in other embodiments, the APPS 108 is located remotely, on a personal computer (PC) (not shown) for example, which can be connected to the printer 102. The APPS 108 includes a feature extraction capabilities and formatting capabilities. An audio file enters the APPS 108 as input and feature extraction techniques are applied to generate a representation 120 of the multimedia content (i.e., a representation of audio content in waveform). The representation or document 120 can include markers for particular features recognized within the multimedia content during feature extraction. For example, the representation 120 could include markers for each time, along an audio timeline, that applause occurs or for each time there is a saxophone solo within a music track. The feature extraction techniques applied may be user defined, or may alternatively be set by a default printer 102 setting. The formatting functionality of the APPS 108 uses the feature extraction results and applies the formatting according to a document format specification (DFS) 104.

**[0034]** In some embodiments, the user can set formatting preferences with regard to the document 120 produced by entering information into fields provided in the DFS 104. In some embodiments, the user can set preferences as to document format and layout, font type and size, information displayed in each line, information displayed in a header, size and location of schedule columns, font colors, line spacing, number of words per line, bolding and capitalization techniques, language in which the document is

printed, paper size, paper type, and the like. For example, the user might choose to have a multimedia document that includes a header in large, bold font showing the name of the multimedia content being displayed (e.g., CNN News segment), and the user can choose the arrangement of the graphical representation of multimedia content displayed per page.

**[0035]** The DFS 104 determines the feature extraction that is applied to the audio data and the format guidelines used to product the output document 120. The DFS 104 is a data structure that can be supplied by an external application, such as a print driver dialog interface on a PC (not shown), or it can be determined internally by interacting with the APPS 108 on the printer's console (not shown). The DFS 104 represents the transformation(s) of the multimedia data. The DFS 104 is used to populate a user interface that is displayed to the user, giving the user formatting options. The DFS 104 determines the feature extraction options presented to the user, which can be applied to the multimedia data. The DFS 104 also determines the format guidelines used to produce the output document.

**[0036]** The DFS 104 can include meta data information about an audio file, such as information about the title of the audio content, the composer of the audio content, and the like. The DFS 104 can also include other information, such as beginning and ending times of a segment (e.g., beginning and ending times of an audio recording), and a specification for a graphical representation of the multimedia data that can be displayed along a time line (e.g., a waveform showing the amplitude of an audio signal over time). The DFS 104 can further include a specification for time stamp markers and meta-data for each time stamp (i.e., a barcode, an RFID tag, a URL, or some other indication for the location where the multimedia data can be retrieved from) that could be displayed along

the timeline, and layout parameters that determine the appearance of the physical multimedia document 120.

**[0037]** In the embodiment shown in Figure 1, the printer 102 additionally comprises processing logic 106 for a printer console and for a print driver interface. The processing logic 106 interacts with the user through a print driver dialog interface (not shown). For example, the processing logic 106 manages the display of a user interface that allows the user to control certain printer actions, such as the processing of the multimedia content or the format in which the multimedia content will be displayed in a multimedia representation 120. Alternatively, the functionality of the user interface can be provided by a web interface, allowing the user to manage printer actions, such as formatting issues, through this web interface. Additionally, the processing logic 106 can return a paper or electronic format for the audio paper. For example, in some embodiments, the user can choose the format in which the representation will be printed. In other embodiments, the printer 102 automatically applies a default setting regarding the format of the representation.

**[0038]** The multimedia document 120 generated by the printer 102 can comprise various formats. For example, the multimedia document 120 can be a paper document, such as an audio paper document 120 of the form shown in Figure 1. The multimedia document 120 produced by the printer 102 can be also stored on digital media. The digital media writing hardware can include, for example, a network interface card, a DVD writer, a secure digital (SD) writer, a CD writer, and the like. The multimedia content can be stored on digital media, such as flash media, a DVD, a CD, and the like.

**[0039]** The multimedia document 120 can have a number of different types of layouts and can display various types of information. Figure 1 provides an example of an audio paper document 120 displaying audio content, though in other embodiments the document may be a video paper document displaying video content. More information about generation of video paper documents is provided in the Video Paper Applications, each of which is hereby incorporated by reference in its entirety.

**[0040]** In the Figure 1 example, the audio paper document 120 includes an audio waveform 112 display of audio information. The layout and format information may specify the length of audio content to be extracted from an audio recording, the arrangement of the audio waveform 112 on the medium, and other like information. For audio information, the printer 102 can extract segments that capture salient features of the audio (or frames that are informative) for a particular segment of the multimedia information. Additionally, as discussed previously, the printer 102 may include feature extraction capabilities (e.g., audio event detection, and the like), allowing the user to search within an audio segment for items of interest, such as for certain speakers, for music, for laughing or yelling, etc. The document 120 produced can display one audio waveform 112 or can divide audio content to be displayed over more than one audio waveform 112. The audio waveform 112 in Figure 1 is displayed vertically, but in other embodiments the audio waveform 112 can be displayed in other arrangements.

**[0041]** Additionally, the audio waveform 112 of Figure 1 includes time stamp markers 114 marking the beginning and the end of the audio content displayed over the audio waveform 112. As an alternative, the audio waveform 112 can include numerous

time stamp markers 114 along the length (i.e., possibly user-defined locations of markers), or the document 120 can include no time stamp markers 114 at all.

**[0042]** In the Figure 1 embodiment of the audio paper 120, the document 120 can contain a header 110. The header 110 provides general information about the audio content included in the document 120. For example, the header 110 may include information about the type of audio content displayed on the document 120 (e.g., “Meeting”), the date of recording of the audio content (e.g., “Nov. 21, 2003”), and the location at which the audio content was recorded (e.g., “RII Conference Room”).

**[0043]** In another embodiment of the present invention, user-selectable identifiers 116 (e.g., a barcode or textual tag) are associated the audio waveform 112. In the Figure 1 example, the user selectable identifiers 116 are displayed on the right side of the audio waveform 112 at user-defined locations, but these can alternatively be displayed anywhere on the page. These identifiers 116 act as index markers, allowing a user to access the associated audio content. For example, in a document 120 printed on paper, the user can physically scan a barcode identifier 116 on the page, and this identifier will point to an audio segment within the audio content displayed on the audio waveform 112. A user selects the user-selectable identifier 116 by scanning the appropriate barcode on the paper document 120 using any type of device (not shown) that has a barcode scanner incorporated into it, such as a cell phone or a personal digital assistant (PDA).

**[0044]** The audio file can be played on a device that allows the random access technique (e.g., barcode scanning) assumed when the document was generated. For example, a document that contains barcodes can be played on a cell phone with a barcode reader and software that can convert barcodes to commands that play an audio file



starting at a given point. Thus, the user-selectable identifiers 116 act as an interface to permit users to access or retrieve the multimedia content displayed on the multimedia document 120.

[0045] As one example, by scanning the barcode of Figure 1, the user can cause the audio segment to begin playing from the marked location on a display device (e.g., a television, a PC monitor, a cell phone screen, a PDA, and the like) and the user can listen to the content. The multimedia document 120 can even provide tactile feedback, by causing a PDA to hum, for example, during parts of a recording that is being played. As another example, the paper multimedia document 120 can also or alternatively include numerical identifiers included instead of or in addition to barcode markers, and the user can type these numerals into a keypad or touchpad (not shown) on the printer 102 or on an external device to direct the system 100 to play an audio segment on a printer display or on the display device. Alternatively, if the audio paper document 120 shown in Figure 1 were in digital format, the system 100 could be configured so that a user could select an audio segment to be played directly from the digital document (i.e., by clicking on the location in the audio waveform 112 with a mouse or other selection device or by selecting a play button).

[0046] The printer 102 is capable of retrieving multimedia information corresponding to the user-selectable identifiers 116. The signal communicated to the printer 102 from the selection device (i.e., device with barcode scanner or keypad for entering in numerical identifiers) may identify the audio segment selected by the user, the location of the audio content to be played, the multimedia paper documents from which the segments are to be selected, information related to preferences and/or one or more

multimedia display devices (e.g., a cell phone) selected by the user, and other like information to facilitate retrieval of the requested multimedia information. For example, the system 100 can access an audio file stored on a PC (not shown), and the system can play this audio content on the user's command.

[0047] The example of Figure 1 further shows text information 118 next to marked locations along the audio waveform 112 in the document 120. In this example, the text information 118 includes portions of a transcript of a conversation that correspond to the marked location along the audio waveform 112. Thus, by selecting the user-selectable identifier 116, the user can cause the audio content to begin playing at the start of the text information 118 that corresponds to the user-selectable identifier 116. Various other types of text information 118 can be also or alternatively displayed along the audio waveform 112 timeline in the document 120, such as summaries of conversations, speaker names, and the like.

[0048] The multimedia document 120 produced by system 100 can be used in a number of different ways. For example, the document 120 provides the user with a convenient way to visually review audio data by searching for particular audio content of interest, providing markers and even text regarding this selected content, and even providing an interface through which the user can access and play audio content. There can also be numerous variations on this type of multimedia document 120. For example, the user can print double-sided video or audio paper. In this example, the user prints a multimedia document 120 on a printer that can apply ink to both sides of a document. The original audio or video paper format can be printed on the front side of the document. The reverse side can show a two-dimensional barcode representation for the

data represented on the front. This format provides a stand-alone paper-based representation that could be stored in a filing cabinet, for example, and subsequent retrieval of the multimedia content would not require reference to an off-paper representation. In the case of double-sided video paper, the video would need to be super-compressed because of the limited capacity of a typical two-dimensional barcode. A combination technology could be used that would extract a rough approximation of the digital data (e.g., the low frequency components of the FFT) from the images printed on the front of the document and supplement that with the higher frequency components, as encoded in the two-dimensional barcode.

**[0049]** As another example, a user can create a perforated multimedia document 120, such as perforated video or audio paper. For example, the user can print a video file that has been segmented into scenes that are each printed on a different perforated strip of paper. Each strip can contain at least one video frame from the video content, and at least one barcode that refers to an online repository of the video data. The strips could be pasted into a notebook or tacked onto a bulletin board, for example. In the case of perforated audio paper, the user can print an audio file that has been segmented by speaker, sound localization, audio event detection, and the like, and each of these segmentation types can be printed on a different perforated strip of paper. For example, one strip could contain barcodes that point to the instances when people were arguing during a meeting. Each strip can contain at least one barcode that refers to an online repository of the audio data. However, because the amount of multimedia data can be limited, a two-dimensional barcode could be used to provide a complete stand-alone representation for the multimedia. These strips could be cut out and easily moved around

by someone who needs to edit the audio recording or by someone who needs to remember only small pieces of the recording. As stated above, the strips could also be pasted into a notebook or tacked onto a bulletin board.

[0050] As another example, the user can create a DVD or CD cover sheet using a multimedia document 120. In this example, the user can print a DVD or CD using this printing technology. Additionally, the printer 102 can be programmed to automatically produce a cover sheet that shows video frames from the scenes segmented from the video file and barcodes that refer to those scenes. This cover sheet can be printed on small paper stock that could be inserted into a special tray in the printer 102, for example. Alternatively, the cover sheet can be printed on normal paper stock and provided with fold-marks that indicate how the paper should be folded so that it fits in the typical DVD holder. A similar cover sheet can be printed for a music CD displaying an audio waveform 122 timeline showing markers for user-selected content and barcodes that refer to the marked portions of the audio content. More information about generating printable representations of multimedia information is provided in the Video Paper Applications, referenced above.

[0051] Referring now to Figure 2, there is shown the architecture of an embodiment of the present invention. In this embodiment, the system 200 includes an APPS 108 that can process audio files that are input into the system 200. The APPS 108 can be located on printer 102 or it can be located on a data processing system (not shown), which could include a PC, a portable computer, a workstation, a computer terminal, a network computer, a mainframe, a kiosk, a standard remote control, a PDA, a game controller, a communication device such as a cell phone, an application server, or

any other data system. Alternatively, the APPS 108 might be located on a printer 102 that is coupled to a data processing system.

**[0052]** In the example of Figure 2, the APPS 108 comprises the following components: a feature extraction module 202 and a formatting module 204. As described previously, the system 200 accesses or receives multimedia information, such as an audio file. The file can be stored on the system 200 or stored on a data processing system (not shown), which is coupled to the printer. In the Figure 2 embodiment, the user can listen to an audio file using any one of various standard multimedia playing tools that allow the user to play back, store, index, edit, or manipulate multimedia information. Examples include proprietary or customized multimedia players (e.g., RealPlayer™ provided by RealNetworks, Microsoft Windows Media Player provided by Microsoft Corporation, QuickTime™ Player provided by Apple Corporation, Shockwave multimedia player, and others), video players, televisions, PDAs, or the like.

**[0053]** An audio file can enter the APPS 108 through a data port 206. This port can include any type of data port, such as an Ethernet connection, over which data can enter printer 102. Additionally, the DFS 104 is input into APPS 108 over connection 208, which couples the APPS 108 to the storage location (not shown) of the DFS 104. Both the feature extraction module 202 and the formatting module 204 can use the DFS 104 information. The DFS 104 defines the feature extraction techniques to be applied to the multimedia content by the feature extraction module 202, and the DFS 104 defines the document formatting information to be used by the formatting module 204

**[0054]** The DFS 104 includes various different types of information. The DFS 104 includes meta data about an audio file for which a representation is being generated.

For example, the DFS 104 can include information such as the title of the audio recording, the artist, the publisher, and the like. The DFS 104 can include beginning and ending times relative to the recording. The DFS 104 can also include a specification for a graphical representation of the audio data that can be displayed along a timeline. For example, the graphical representation can be an audio waveform, as discussed in Figure 1. The waveform can show the amplitude of the audio signal over time, and the user can zoom into and out of the audio waveform when necessary. Another example would be a JPEG for a waveform. The DFS 104 can also include a specification for time stamp markers and meta data for each time stamp or user-selectable identifiers (i.e., textual tags or barcodes), that could be displayed along the timeline.

[0055]       Layout parameters can also be defined in the DFS 104, in which the parameters determine the appearance of the physical document 120 created. The layout parameters can include, for example, a specification for the portion of the timeline that will be displayed on each page of document 120. The generation of the layout can be determined by a default behavior specification, stored in the printer default settings (e.g., Printer Properties). This can include the autonomous productions of a paper document 120 or an interactive process using a user interface on a printer's console, a web page, etc.

[0056]       The feature extraction module 202 produces the graphical representation, and the user-selectable identifiers 116 and time stamps specified in the DFS 104. Examples of a graphical representation include a curve that shows the amplitude of an audio file over time. Examples of other features that could be used to produce user-

selectable identifiers 116 include the detection of solos in a musical performance, speech recognition, applause detection, detection of music, and the like.

**[0057]** The formatting module 204 is coupled to the feature extraction module by connection 210. Feature extraction data is sent over connection 210 to the formatting module 204 for use in formatting the document 120. The formatting module 204 converts the audio features and the DFS 104 into a document representation that can be rendered on paper or as an electronic file, such as a PDF document. The DFS 104 contains detailed information about the fonts to be used and other information that is typically provided by a document-formatting package (e.g., Microsoft Word). This layout information will be included in the “layout” field of the DFS 104, discussed below.

**[0058]** The system 200 of Figure 2 can also include a processor (not shown), in printer 102, which processes data signals. The processor (not shown) may comprise various computing architectures including a complex instruction set computer (CISC) architecture, a reduced instruction set computer (RISC) architecture, or an architecture implementing a combination of instruction sets. The system 200 can include a single processor or multiple processors. Main memory (not shown) may store instructions and/or data that may be executed by processor 214, including the software and other components of system 200. The instructions and/or data may comprise code for performing any and/or all of the techniques described herein. Main memory (not shown) may be a dynamic random access memory (DRAM) device, a static random access memory (SRAM) device, or some other memory device known in the art.

[0059] When printer 102 receives a print request, the request and the associated multimedia data is transferred to a processor (not shown), in some embodiments. The processor interprets the input and activates the appropriate module. In some embodiments, the processor is coupled to and controls the feature extraction module 202 for transforming multimedia content. Additionally, the processor is coupled to the formatting module 204 for controlling formatting of document 120, in some embodiments. The APPS 108 generates the appropriate document-based representation and can interact with the user through a print driver dialog interface (not shown) to modify the parameters of the document 120 generation and to preview the results. The results and parameters of the multimedia transformation are represented in the DFS 104. The processor (not shown) can also manage generation of a document 120, by communicating with and sending print job information to a conventional printer (not shown), and the conventional printer (not shown) generates a paper output. As previously described, the document 120 can also include user-selectable identifiers, such as barcodes, and other links to multimedia data stored by the printer 102 or stored in a specified online database (not shown).

[0060] In operation, the system 200 provides methods for printing multimedia content, and in the specific examples given in the Figures, the system 200 provides methods for printing audio content. Referring now to Figure 3, there is shown a flowchart that describes processing steps in an audio paper production system 108. The APPS 108 is coupled to a control program that runs the subroutine process, as described below. In this embodiment, the processing steps of the APPS 108 include inputting 302 an audio file into the system and inputting 302 the DFS 104 into the system. Based on



the user's instructions, the APPS 108 determines whether or not a graphical representation has been requested. If not, the APPS 108 moves on to determining if feature extraction is required. If so, the APPS 108 then calls 304 the feature extraction module 202 of the system 200 to produce the graphical representation using the audio file information and the information provided in the DFS 104. The APPS 108 updates the DFS 104 by adding 306 a symbolic form or representation of the feature extraction results as one of the document specification fields listed in the DFS 104.

**[0061]** As a next step in the process, the APPS 108 determines if feature extraction is required. If not, the APPS 108 moves onto calling 312 the formatting module 204 to produce the document type specified in the DFS 104 output format listed in the "layout" field of the DFS 104. If so, the APPS 108 calls 308 the feature extraction module 202 to produce the markers requested in the DFS 104 using the audio file information and the information included in the DFS 104. The APPS 108 then adds 310 marker data to the DFS 104. Once this step is completed, the APPS 108 calls 312 the formatting module 204 to produce the document type specified in the DFS 104 output format listed in the "layout" field of the DFS 104.

**[0062]** Referring now to Figure 4, there is shown a flowchart describing the operations of the formatting module 204. The formatting module 204 is coupled to a control program that runs the subroutine process, as described below. In this embodiment, the processing steps of the formatting module 204 include inputting 402 the results of feature extraction conducted by the feature extraction module 202. For each page listed in the "layout page" field of the DFS 104, the formatting module 204 determines if the formatting of the page is finished. If it is, the formatting module 204

sends a return message to the control program. If it is not finished, the formatting module 204 formats 404 meta-data in the formatting module 204 as specified in the “layout” field of the DFS 104 regarding “meta-data placement.” The formatting module 204 then formats 406 the graphical representation created by the feature extraction module 202, as specified in the “layout type” field of the DFS 104, based on the result of the feature extraction. The formatting module 204 generates 408 barcodes, according to the “marker type” field of the DFS 104, the “marker frequency” field of the DFS 104, and the “marker n” field of the DFS 104. The markers are then formatted 410 as specified in the formatting module 204 given the barcodes. The system then renders 412 the page given the formatted meta-data, graphical representation, and markers. Once this process is finished for a page, the formatting module 204 then runs through this process for the next page in the “layout pages” field of the DFS 104, and for all other pages, until all pages in the “layout pages” field have been formatted.

[0063] Referring now to Figure 5, there is shown a flowchart describing generation of barcodes for multimedia documents 120. The APPS 108 is coupled to a control program that runs the subroutine process, as described below. In this embodiment, the processing steps include inputting 502 information, including barcode type (e.g., interleaved 2 of 5), number of identifier digits in barcode, number of time stamp digits in a barcode, and time stamp value. The system then reads 504 the identifier field from the formatting module 204, and then converts 506 the identifier into a right-justified decimal string. The system then determines if the length of the right-justified identifier is greater than the number of identifier digits allowed in the barcode. If so, then the system returns an error code to the control program. If not, then the system converts

508 the time stamp value into a right-justified decimal string. The system then determines if the length of the right-justified time stamp is greater than the number of time stamp digits allowed in the barcode. If so, then the system returns an error code. If not, then the system appends 510 the right-justified time stamp, which is padded on the left with zeros, to the right-justified identifier. The system then renders 512 a barcode image of the specified type, containing identifier information plus time stamp information. The system sends a return message to the control program signaling the end of the operation.

[0064]        Though the above-described flowcharts are discussed in reference to audio content, these methods can also apply to video or other media content. The figures that follow show examples of the results of applying different combinations of format specification, feature extraction, and parameters for the audio paper generation algorithm. As stated previously, the format specifications, feature extraction and parameters can also be used to generate a document displaying another type of media content.

[0065]        Figure 6 shows a graphical representation of audio paper with regularly spaced user-selectable identifiers 116 and the associated DFS 104 with audio feature extraction specification 602 for creating the audio paper. In Figure 6a, there is shown a DFS 104 for specifying the layout and content for generating an audio paper document 120. It includes various DFS 104 fields, in which information regarding layout and content is specified. In this example, there is a “type” field 604, listing the type of audio content included in the document 120 (e.g., a musical recording). The “identifier” field 608 lists identifying information that is to be included in the barcode or user-selectable identifier 116. The “title” field 610 lists the title of the musical recording (e.g.,

Locomotion). The “artist” field 612 lists the name of the artist who created the audio content (e.g., John Coltrane). The DFS 104 includes a “collection” field 614, specifying the music collection or album in which the recording is included (e.g., Blue Train). The DFS 104 also includes a “publisher” field 616 and “publication date” field 618 that specify who published the recording and on what date (e.g., Blue Note Records, in 1957). The “begin time” field 620 and the “end time” field 622 list the time at which the audio content begins (e.g., “00:00:00”) and the time at which the content ends (e.g., “00:07:14”). The “graphical representation” field 624 describes the type of graphical representation of audio content that will be included in the document 120 (e.g., an amplitude curve).

**[0066]** The DFS 104 also includes information about user-selectable identifiers 116, or markers, to be included in the document 120, along with the layout specifics of the document 120. There is shown a “marker type” field 628 and a “marker frequency” field 630, which specify the type of marker to be included in the document 120 (e.g., a barcode), and the frequency at which the marker should appear in along the graphical representation (e.g., at 30 second intervals). Additionally, the layout fields give information about the layout of the document 120. In Figure 6a, there is shown a “layout type” field 632 that specifies the arrangement of the contents of the audio paper. For example, the layout type can include one horizontal timeline to be displayed on the document 120, or it could instead include two vertical timelines. The “layout pages” field 634 specifies the number of pages of the document 120. The “layout marker placement” field 636 specifies the location at which user-selectable identifiers 116 or markers should be displayed (e.g., above the graphical representation). Additionally, the

“layout meta-data placement” field 638 lists information about the placement of meta-data on the document 120. The meta-data can include a header 110 or other meta-data.

**[0067]** The DFS 104 of Figure 6a displays just one example of a set of information about the media representation being generated. In other embodiments, the DFS 104 can include a number of other fields including, but not limited to, a field for picture information, for hypertext links, for biographies of artists, for birth/death dates of artists, for address information of artists, for information on where to purchase the displayed media content (i.e., link to website for purchasing the album), and the like. Some other examples of DFS 104 fields are also discussed below. This is a non-exhaustive list of variations, and numerous other types of information could be incorporated.

**[0068]** Also shown in Figure 6a is the audio feature extraction specification 602. The feature field 606 defines the feature extraction applied to the audio content. In this example, the audio feature extraction 602 is an audio amplitude extraction and graphical approximation. Thus, the document 120 shows an audio waveform 112. In this example, an SVG file is output.

**[0069]** In Figure 6b, there is shown a graphical representation of an audio paper document 120, according to one embodiment of the present invention. In this document 120, there is shown a header 110, with header information according to the specifications in the DFS 104. The header 110 is also positioned as specified in the DFS 104 (i.e., in this case, it is centered at the top of the page). The document 120 displays an audio waveform 112 or amplitude curve along a timeline. In other embodiments of the present invention, the timeline can be represented by a single straight line, or other variations of a

graphical representation. The timeline runs from “00:00:00” to “00:07:14,” which corresponds to the length of the audio recording. Time stamps 114 are shown at three places along the audio waveform 112, marking the beginning time and ending time of the recording, along with a time stamp 114 marking a location in the middle of the section. The document 120 can show more than three time stamps 114 or no time stamps 114 at all, according to the user’s preferences.

**[0070]** In addition, the document 120 displays user-selectable identifiers 116 (e.g., barcodes), which provide an interface by which the user can access the audio content at locations along the timeline. In some embodiments, the user can specify particular locations for each user-selectable identifier 116. In this example, the user has specified that the document 120 should include barcode markers at every 30 seconds along the timeline. These user-selectable identifiers 116 are displayed in a “stair-step” fashion, rather than in one long line, to allow for easier selection of each individual identifier 116. However, the arrangement of user-selectable identifiers 116 can vary greatly, and can be specified in the DFS 104. As described previously, the user can select any user-selectable identifier 116 in the printed document to play the associated audio content. For example, the user could scan a barcode, using a cell phone with a barcode scanner, at any location along the timeline to play the recording starting from that point on the cell phone screen or other display device.

**[0071]** Referring now to Figure 7, there is shown a graphical representation of a document 120 with user-selectable identifiers 116 for each musical solo in the recording, and the associated DFS 104 with audio feature extraction specification 602 for creating the audio paper 120. The DFS 104 shown in Figure 7a is similar to that shown in Figure

6a, but the example of Figure 7a includes a few variations. The Figure 7a example of a DFS 104 includes a “feature extraction” field 702 that lists the feature extraction to be applied to the audio content. In this case, the feature extraction includes marking of musical solos within the audio content, in which the output shows the instrument name and the time when the solo began. In the example of Figure 7a, there is shown a “marker type 1” field 704 and a “marker type 2” field 706, and these fields specify two types of user-selectable identifiers 116 to be displayed in the document 120. For example, the document 120 will include a marker type 1 that displays an instrument name that is shown above a barcode that is shown above a time stamp 114. In the example, a marker type 2 will be a barcode (i.e., a second barcode displayed at defined locations under the graphical representation). The DFS 104 also includes a “layout marker 1 placement” field 710 and a “layout marker 2 placement” field 712. These fields specify where each marker will be shown on the document 120, such as under the timeline or above the timeline.

**[0072]** Figure 7a also shows an audio feature extraction specification 602 in the “feature” field 606. The audio feature extraction specification 602 in this example includes audio amplitude extraction and graphical approximation, along with an SVG file output. Additionally, the audio feature extraction specification 602 includes a musical solo extraction, which outputs the beginning times of each solo and the instrument used in each solo. In this example, one type of feature extraction is being applied to audio content. However, the system can be configured to apply any number of feature extraction types at a time. Other examples of feature extraction include, but are not limited to, speech detection, speaker detection, speaker recognition, video/audio event

detection, video foreground/background segmentation, face detection, face image matching, face recognition, face cataloging, video text localization, optical character recognition (OCR), language translation, frame classification, clip classification, image stitching, audio reformatter, audio waveform matching, audio-caption alignment, video OCR and caption alignment, sound localization, radio transmission recognition, audio/video range selection by a slider, speaker segmentation, profile analysis, color histogram analysis, clustering, motion analysis, distance estimation, scene segmentation, license plate or automobile recognition or motion analysis, and the like. This is a non-exhaustive list of variations, and numerous other types of extraction could be incorporated in the present invention.

**[0073]** Figure 7b shows a graphical representation of a document 120 according to the DFS 104 and audio feature extraction specification 602 of Figure 7a. The document 120 includes a header 110, an audio waveform 112 displayed horizontally, and time stamps 114 at locations along the bottom of the audio waveform 112. In this example, there are also user-selectable identifiers 116 included near each timestamp 114. The user can select these user-selectable identifiers (e.g., scan the barcode) to begin playing the audio content at the location of the marker. For example, if the user scans the barcode shown above time “00:00:00,” the recording will start playing from the beginning. Since a solo extraction was applied to the audio content displayed in document 120, the audio waveform 112 includes markers for each musical solo. The markers include the text description 714 that explains the type of solo being marked (e.g., a saxophone solo), a user-selectable identifier 116 that provides an interface to the solo, and a time stamp 114 showing the location of the solo within the audio content. For



example, by scanning the barcode under the “sax” solo in the printed document, the saxophone solo will start playing from the beginning on a display device.

[0074] Referring now to Figure 8, there is shown a graphical representation of a document 120 showing a timeline for a radio program, and the associated DFS 104 with audio feature extraction specification 602 for creating the audio paper 120. The DFS 104 shown in Figure 8a is similar to that shown in Figure 6a, but the example of Figure 8a includes a few variations. The example of Figure 8a shows the DFS 104 for a radio program, and the DFS 104 includes an “annotation” field 802 that adds an extra annotation to the document 120 regarding the radio program. In this example, the annotation shows that the guest on the program is “Bill O’Reilly.” Thus, a summary of available meta information about a radio talk show, such as the time it occurred, the name of the host, and its duration can be printed on paper together with barcodes for each portion of the conversation and indications of when commercial breaks occurred. The names of the participants could be included if they are known. The barcodes could point to audio data recorded separately by the user of the system, or they could point to audio data on a web site provided by the talk show. This could be coupled to software that post-processes the recording, producing the document and the streamable media file as well as a web page. Further utility would be provided by actively linking the production process to annotation performed online while the talk show occurs. This annotation can be performed at the radio station while the show is happening since the producers would have access to information not available to a listener, such as the phone numbers of people calling the program.

[0075] Figure 8b shows a document 120 displayed according to DFS 104 with audio feature extraction specification 602. The document 120 includes a header 110 showing the title, date, and annotation information for the radio program, an audio waveform 112, time stamps 114, and user-selectable identifiers 116 displayed in a “3-step staircase” fashion.

[0076] Referring now to Figure 9, there is shown a graphical representation of a document 120 showing a timeline for a radio program with markers for keywords, and the associated DFS 104 with audio feature extraction specification 602 for creating the audio paper 120. The DFS 104 shown in Figure 9a is similar to that shown in Figure 8a, but the example of Figure 9a includes variations on the marker information. The “marker type” field 628 of Figure 9a shows a marker type that includes keywords, barcodes, and timestamps. The “marker frequency” field 630 shows that the frequency is “user-defined.” Thus, in this example, the user has selected each marker to be displayed along the timeline. In the “marker” fields 902, the user has made selections for markers 1 through 11. For marker 1, for example, the user has defined the marker to include a barcode, a timestamp, and text describing the marker (e.g., “WTC”), and likely describing the audio content that is being marked. The user has also defined the vertical position of each marker, as “vert. pos. 1,” “vert. pos. 2,” or “vert. pos. 3,” along the timeline. These positioning specifications determine where the marker will be positioned vertically, chosen from a number of stair-step positions above the timeline. The audio feature extraction 602 is again an audio amplitude extraction with graphical approximation.

[0077] Figure 9b shows document 120 displayed according to the DFS 104 and audio feature extraction specification 602 of Figure 9a. The document 120 includes a header 110 for the radio program, an audio waveform 112, time stamps 114 under the audio waveform 112, and markers displayed at user-defined positions along the timeline. The markers include text information 714 set by the user to describe the content in the radio program that is being marked. Additionally, the markers include a user-selectable identifier 116 and a timestamp 114.

[0078] Referring now to Figure 10, there is shown a graphical representation of a document 120 showing a timeline for a radio program with audio feature extraction for search terms. The DFS 104 of Figure 10a shows, in the “feature extraction” field 702 that speech recognition and keyword match techniques were applied to the audio content. In this example, the user searched for “New York Times” or “fair and balanced” as search terms. Figure 10a shows the “marker type” field 628 that includes a matching search term, barcodes, and timestamps. The “marker frequency” field 630 shows that the frequency is “user-defined.” Thus, in this example, the user has selected each marker to be displayed along the timeline. In the “marker” fields 902, the user has made selections for markers 1 through 11. For marker 1, for example, the user has defined the marker to include a barcode, a timestamp, and text describing the marker (e.g., “fair and balanced”), and the vertical position of each marker.

[0079] The audio feature extraction specification 602 again includes an audio amplitude extraction with graphical approximation. The audio feature extraction specification 602 also includes speech recognition, along with term matching to a given list of keywords. Thus, the user searched for locations within a radio program in which

the speakers use particular terms, and these locations are marked along the timeline, possibly along with a transcript of a portion of or all of the speech. As an alternative, the user could apply speech recognition alone, recognizing any point in the audio content in which speech occurred. Since the speech recognition output may be noisy, some representation for the confidence of the recognition can also be included, so the user can see which words or phrases are more likely to be correct. For example, the document 120 could include colors or variations in font size to represent recognition confidence. The highest confidence decisions could be shown in red, 12-point font, while the lowest confidence decisions could be in blue, 8-point font. User-selectable identifiers 116 can be included for each decision or for only the ones with the highest confidence.

[0080] Other examples of audio feature extractions that could be applied to audio content include speaker detection and speaker recognition. The speaker detection extraction can recognize a group of equivalent speakers in a recording and determine when the same person was speaking. This can be represented along a timeline by segments annotated with a limited palette of colors, showing a different color for each speaker, and the same color for the same speaker. This might be used for scanning through a long recording looking for only the comments of a specific person. Speaker recognition extraction identifies the actual people who spoke during an audio recording. The symbolic identity of the people can be computed and added next to segments of the timeline together with barcodes that when swiped would play the audio from the beginning of that segment. This would allow one to scan the printout and see who participated in a meeting. An alternative version could print a list of names and could place bar codes next to those names. A user could swipe those barcodes and listen to the

parts of the recording when that person spoke. A further example would retrieve face images for those people and print them next to the names and barcodes. The audio data could also be embedded in a two-dimensional bar code, thus providing a completely stand-alone representation for the audio file.

**[0081]** Figure 10b shows document 120 displayed according to the DFS 104 with audio feature extraction specification 602 of Figure 10a. The document 120 includes a header 110 for the radio program that also shows an annotation of “keyword search terms,” an audio waveform 112, time stamps 114 under the audio waveform 112, and markers displayed at user-defined positions along the timeline. The markers include text information 714 set by the user to describe the content in the radio program that is being marked. In this case, the text information 714 is the specific term that was found in the audio content. Additionally, the markers include a user-selectable identifier 116 and a timestamp 114. By selecting the user-selectable identifier 116, the user can hear the audio content in which the search term was used.

**[0082]** In Figure 11, there is shown a graphical representation of a document 120 showing a timeline for a radio program with audio feature extraction for applause events. The DFS 104 of Figure 11a shows, in the “feature extraction” field 702 that applause detection was applied to the audio content. The audio feature extraction specification 602 includes applause detection timestamps. Thus, the user searched for locations within a radio program that applause events occurred, and these locations are marked along the timeline.

**[0083]** Figure 11b shows document 120 displayed according to the DFS 104 and audio feature extraction specification 602 of Figure 11a. The document 120 includes a

header 110 for the radio program that also shows an annotation of “applause events shown,” an audio waveform 112, time stamps 114 under the audio waveform 112, and markers displayed at user-defined positions along the timeline. As an alternative, the system could mark when other events occurred in the audio content, such as laughter, loud conversations, doors slamming, etc. This could be used to quickly scan a recording of a meeting, for example, for points at which arguments occurred in the meeting. The markers include a user-selectable identifier 116 and a timestamp 114. By selecting the user-selectable identifier 116, the user can hear the audio content in which the applause occurred. The user-selectable identifier can refer to audio data stored off-line or to audio embedded in two-dimensional identifiers (e.g., barcodes) printed on the paper document.

[0084] Sound localization techniques could also be applied to audio content. In this example, a timeline representation can include directional indicators that point to places in a room where the recording was done. This allows users to quickly scan the timeline and determine when, for example, the person in the southeast corner of the room or the person across the table from them, was speaking. This can be applied, for example, with fixed installations that have multiple-microphone setups that can be calibrated to perform sound localization. It can also be used with appropriately equipped portable recorders, such as those used by professionals who record interviews.

[0085] In Figure 12, there is shown a graphical representation of a document 120 showing a timeline for a radio program with audio feature extraction for music events. The DFS 104 of Figure 12a shows, in the “feature extraction” field 702 that music detection was applied to the audio content. The audio feature extraction specification 602 includes music detection timestamps. Thus, the user searched for locations within a radio

program that music events occurred, and these locations are marked along the timeline. Additionally, the “layout type” field 632 shows the layout type to include two vertical timelines that are split in half.

[0086] Figure 12b shows document 120 displayed according to the DFS 104 and audio feature extraction specification 602 of Figure 12a. The document 120 includes a header 110 for the radio program that also shows an annotation of “music events shown,” an audio waveform 112, time stamps 114 near the audio waveform 112, and markers displayed at user-defined positions along the timeline. The markers include a user-selectable identifier 116 and a timestamp 114. By selecting the user-selectable identifier 116, the user can hear the audio content in which the applause occurred. The timelines are displayed according to the layout type shown in the DFS 104, with the timeline split into two halves, and each half displayed vertically.

[0087] Multimedia paper can also be used for generating representations of voice mail messages. A user can generate a summary of the available meta information about a collection of voice mail messages, such as the phone number of the calling party, the result of looking up that phone number in an internet search engine (which can often show the name of the caller, their address, and a map showing their location), as well as the date, time, and duration of messages. Each block of meta information could be printed next to a barcode that would retrieve the audio information from a remote network location or it could be represented in a two-dimensional barcode that could be played directly from the paper, thus obviating the need for any off-device access. The paper document provides value to users by providing extra information that can be retrieved and added to the paper document (e.g., internet search engine information).

Also, the paper document itself would provide users with the ability to write notes about the voice mail messages on the document.

**[0088]** Multimedia paper can also be used in generating representations of public safety radio transmissions. A summary of the available meta information about the recording of one or more public safety (e.g., police, fire, etc.) radio transmissions, including the date, time, duration, car number, officer name (if available), can be printed on paper together with barcodes that reference an off-line representation for those recordings. Two-dimensional bar codes can also be used that directly encode audio data. This provides a stand-alone representation that can be used independently of a network connection. The meta information can be computed by signal processing algorithms applied to the recorded audio, or it could be computed from digital side channel information provided in the radio transmissions (e.g., Motorola digital radio information). Alternatively, it could be provided digitally at the radio dispatcher's console. This system could assist managers who need to selectively inspect the radio logs, or it could assist members of the public who want to observe public safety procedures.

**[0089]** Multimedia paper can also be used in generating representations of aviation radio transmissions. A summary of the available meta information about the recording of one or more aviation radio transmissions, including the date, time, duration, flight name, origin, destination, current position, of a flight when a particular transmission occurred, can be printed on paper together with barcodes that point to an online form of the audio recording. The meta information can be extracted directly from the mode S transponder returns, assuming suitable equipment is available. Additional meta information could be retrieved from various online services that track flight



progress in real-time. Speech recognition applied to the audio recording could provide symbolic information that could be used to compute links to the online data. This would obviate the need for a direct link to the mode S data and would make this system usable by people without access to FAA equipment.

**[0090]** While the present invention has been described with reference to certain preferred embodiments, those skilled in the art will recognize that various modifications may be provided. Variations upon and modifications to the preferred embodiments are provided for by the present invention, which is limited only by the following claims.